25s4 - Protein Simulation IV on 4/14/2025

Objectives of the Lecture

By the end of this lecture, students should be able to:

1. Explain the physics and mathematics underpinning our ability to conduct molecular dynamics simulations
2. Know how to work with different thermodynamic systems in which different parameters are fixed
3. Determine whether molecules experience attractive or repulsive forces given the potential energy of interactions based on separation distances
4. Describe the Lennard-Jones potential both conceptually and mathematically
5. Identify the optimal method of integration used in molecular dynamics and the tradeoffs

Key Concepts and Definitions

- Molecular (or atomic) dynamics: simulation methodology to understand the dynamics of molecules (or atoms) over time. It operates at all length/time scales, but is applied herein on molecules
- Coupled 1st order differential equations: system of two or more simultaneous differential equations where rate of change of each variable depends on itself and other variables in the system
    - $\frac{d\vec{r}_i}{dt} = \vec{v}_i$: time rate of change of position is equal to velocity
    - $\frac{d\vec{v}_i}{dt} = \frac{\vec{F}_i}{m}$: time rate of change of the velocity is acceleration, which is also the sum of forces divided by mass
    - here, i = 1, …, N atoms/objects/particles
    - here, $\vec{r}_i = (x_i, y_i, z_i)$ and indicates a point in the system
- Energy conserving: energy is constant and does not change with time
- Ensembles: denotes a collection of systems, described by a specific subset of defining thermodynamic variables. Initial conditions finish defining a particular member of an ensemble to a system of interest
    - An example ensemble of interest is the NVE ensemble, where number, volume, and energy is constant but temperature is not
- Packing fraction (Φ): the ratio of the volume occupied by the particles to the total volume in a system

Main Content/Topics

1) **Simulating particles via molecular dynamics under various ensembles**

When no thermostat is used, the simulation is in the NVE ensemble, meaning temperature is not controlled directly. The system evolves by integrating Newton's equations of motion, conserving total energy.

Total energy:
E=K+U — total energy is the sum of kinetic energy (K) and potential energy (U).

Kinetic energy:

$K = \frac{1}{2}\sum_{i}^{N} m_i v_i^2$ — the usual definition over N particles.

Conservation of energy:
Energy is constant over time in a closed system. The derivative of E is 0.

Defining a Thermodynamic System (Ensemble Types):
Ensembles describe systems in which different parameters are fixed:

- NVE (constant Number, Volume, Energy) - Microcanonical Ensemble
  - Least useful, since real systems rarely maintain good separation from their surroundings.
  - Primary focus of the lecture
- NVT (constant Number, Volume, Temperature) - Canonical Ensemble
  - Good for dynamics and energetics
  - Essential in Equilibration
  - Requires a thermostat
- NPT (constant Number, Pressure, Temperature) - Isobaric-Isothermal
  - Best at mimicking essential conditions, phase transitions
  - Requires a barostat as well as a thermostat
- NPE (constant Number, Pressure, Energy) - rarely used
- Instead of opting for a classical ensemble, we can use packing fraction as a stand-in to define volumes

The role of packing fraction in an NVE ensemble:

- Imagine an isolated system of atoms moving around in a two-dimensional box of L x L. The box contains N number of atoms. The temperature T can also be described as K/N, or kinetic energy divided by the number of atoms. Then, the packing fraction (Φ) determines the behavior of the system. This is $\phi = \frac{N\pi r^2}{L^2}$, where r is the radius of the atom.

2) How forces relate to potential energy using vector calculus

Force from potential energy gradient:

$F_i = -\nabla U$— classical relation: force is the negative gradient of potential energy.

Expanded gradient:

As a reminder, the gradient can be expanded to give partial derivatives along each axis.

$$F = \frac{\partial}{\partial x}U_x\hat{x} + \frac{\partial}{\partial y}U_y\hat{y} + \frac{\partial}{\partial z}U_z\hat{z}$$

Potential energy example:

$U = mgz$ — gravitational potential energy (m = mass, g = gravity, z = height).

Force from gravity:

$F_i = -mg\hat{z}$ — points downward along z-direction.

Coordinate system:

We maintain a Cartesian coordinate system in 3D defined by unit vectors, $\hat{x}, \hat{y}, \hat{z}$

- Note the following relations of the orthonormal basis vectors
  - $\hat{x} \cdot \hat{x} = 1$
  - $\hat{x} \cdot \hat{y} = 0$
- This allows for ease when conserving quantities with respect to these units
- Clarification: "hat" on vectors means unit vectors (magnitude = 1).

3) Simulation with n-body and field interactions:

$U(r_i, r_{ij}, \theta_{ijk})$

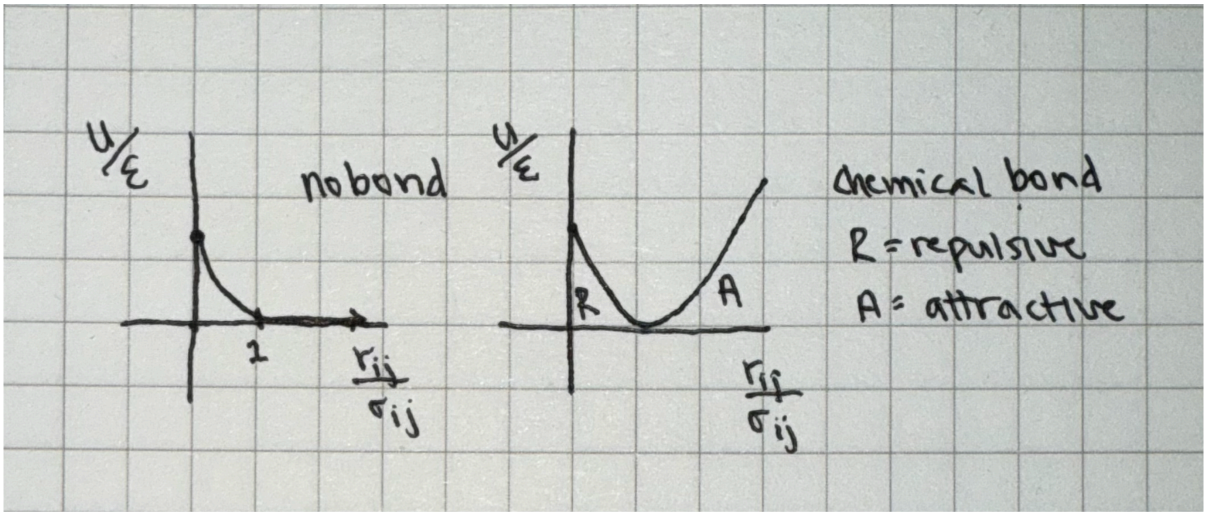
We can interpret potential energy of interactions to be dependent on three factors: r_i, r_{ij}, θ_{ijk}. Here, r_i refers to external fields (magnetic, electric, etc) which we can disregard for now. r_{ij} refers to the separation of contact between the particles of interest, where direction of separation matters. θ_{ijk} refers to three-body interactions and is an advanced topic we can disregard for now. In this class, we consider potential energy to be dependent on r_{ij}.

Now, let’s look at a particular implementation:

$$\frac{U(r_{ij})}{\epsilon} = \frac{1}{2}\left(1 - \frac{r_{ij}}{\sigma_{ij}}\right)^2 \theta\left(1 - \frac{r_{ij}}{\sigma_{ij}}\right)$$

- Only ϵ has units (energy), the rest remains unitless through the introduction of scaling parameters
- r_{ij} and σ_{ij} have the same units (of length) and so do not contribute overall units
- The theta here is a Heaviside step function. For our purposes, it merely serves as a cutoff to exclude interactions beyond our distance parameter

Given this, the potential energy diagram follows as:



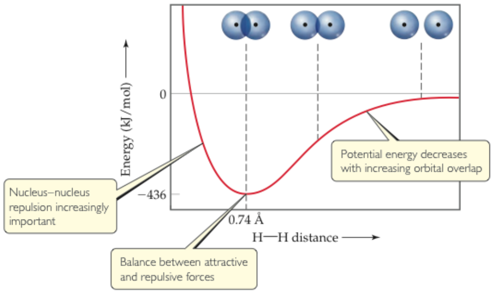
The derivative of the potential energy diagram indicates whether the forces are attractive or repulsive. If the derivative is negative, the forces are repulsive. If the derivative is positive, the forces are attractive. In the figure above, the left graph shows a potential energy diagram for interactions with no bonds, while the right graph shows interactions with chemical bonds. In both graphs, the left side shows repulsive forces.

Recall that $\vec{F}_i =-\frac{dU}{\partial r_{ij}}$. Then:

$$\vec{F}_i =-\frac{dU}{\partial r_{ij}} =-\frac{d}{\partial r_{ij}}\left(\frac{\epsilon}{2}\left(1 - \frac{r_{ij}}{\sigma_{ij}}\right) \theta\left(1 - \frac{r_{ij}}{\sigma_{ij}}\right)\right) = \frac{\epsilon}{\sigma_{ij}}\left(1 - \frac{r_{ij}}{\sigma_{ij}}\right) \hat{r}_{ij}$$

Here, the derivative is negative, which cancels the negative sign to make the term for $\vec{F}_i$ positive, indicating a repulsive force. In the opposite case, if the derivative is positive, $\vec{F}_i$ would be negative, indicating an attractive force.

4) Lennard-Jones potential:



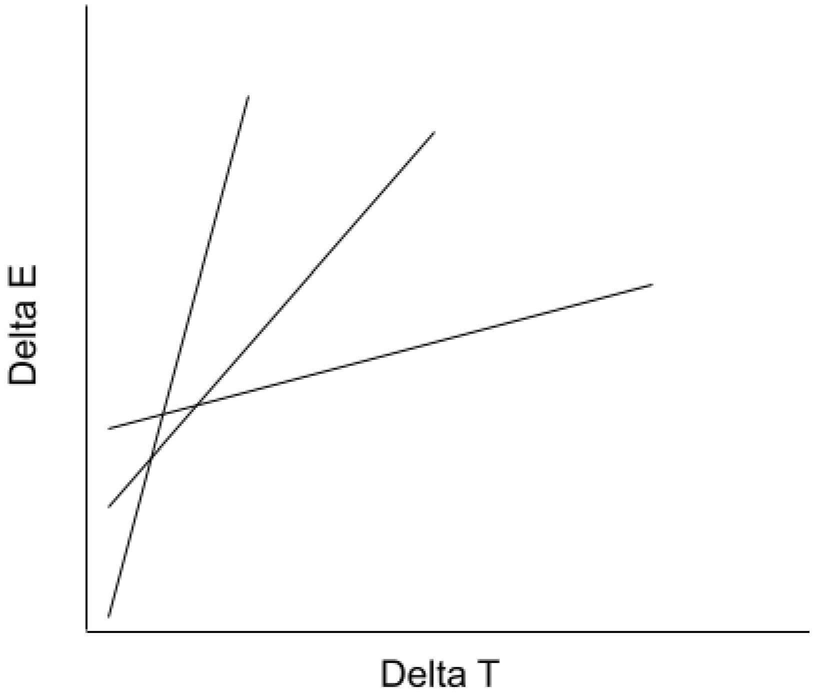
The Lennard-Jones potential can be described as:

$$U(r_{ij}) = 4\epsilon\,[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}]$$

The Lennard Jones potential describes the potential energy of interaction of two non-bonding molecules based on their distance of separation. It consists of a steep repulsive force that goes up to infinity, a minimum, and an attractive dipole-dipole force which goes out to infinity. Intuitively, this makes sense as molecules do not like to sit directly on top of one another (steep repulsive force) and experience dipole-dipole attraction at larger separation distances. This attractive force can pose a problem in molecular dynamics, where a decision must be made to truncate or smoothen the force.

5) Choosing an Integration Regime

Running simulations of differential equations ultimately require methods of integration, which operate at varying degrees of uncertainty. Though the ideal equations maintain energy conservation, integration creates errors dependent on the timestep. Different methods of integration are primarily differentiable by their orders of dependency on the time step. This can be visualized on a log plot between log(delt) and delE which should manifest as shown:



Though higher order methods are most accurate for small timesteps which we may optimally hope to use, they lead to large error at larger timesteps which may be necessary for a sufficiently long simulation. For the purposes of Molecular Dynamics of proteins, Velocity Verlet Integration is considered the best compromise with timesteps of interest, and is a second order method. In it positions and velocities are updated as follows:

$$r_i(t + \Delta t) = r_i(t) + v_i(t)\Delta t + \frac{1}{2}\frac{F_i(t)}{m}\Delta t^2$$

$$v_i(t + \Delta t) = v_i(t) + \frac{\Delta t}{2}(F_i(t) + F_i(t + \Delta t))$$

Discussion/Comments

List all suggested reading here and please answer:

Are the readings for the class useful? If so, are the specific subsections useful or would change. If not, are there other references you could suggest? Please suggest one.

There are no suggested readings for this class, and neither ESL nor ISL reference concepts related to molecular dynamics. However, Suggested References below points to a few resources for learning about and carrying out Molecular Dynamics simulations.

Suggested references for many of the key concepts:

Allen, M. P.; Tildesley, D. J. Molecular Dynamics in *Computer Simulation of Liquids*; Oxford University Press, 2017; pp. 95-146. DOI: 10.1093/oso/9780198803195.001.0001.

For practical applications:

- http://www.mdtutorials.com/gmx/
- https://ambermd.org/tutorials/